# Evaluating the Restrict to MeSH Algorithm

Final Report for NLM Summer Research Rotation
August 3, 2006

Trainee: Michael Bales, M.P.H., Dept. of Biomedical Informatics, Columbia University
Mentor: Olivier Bodenreider, M.D., Ph.D., Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine

*This report contains the preliminary results for the project. Final results will be included in a planned future conference submission (MedInfo 2007).*

## Introduction

The goal of the U.S. National Library of Medicine (NLM) is to collect, organize and make available biomedical literature to advance medical science and improve public health. To this end, the Library has developed and maintains a number of online resources. One of the most widely-used is MEDLINE, an online database of references and abstracts from biomedical journals. The ability of patrons to search MEDLINE depends largely on Medical Subject Headings (MeSH), the NLM's controlled vocabulary thesaurus. MeSH consists of a set of terms arranged into a hierarchical structure. These headings allow for searching the literature at various levels of granularity. MeSH headings are also used for a variety of other purposes, such as for browsing ClinicalTrials.gov, the National Clinical Trials Registry(1).

MeSH has broad coverage of topics in a variety of domains. Despite this broad coverage, there are many concepts which appear in the literature but which are not enumerated in MeSH. These are largely concepts that are at a finer level of specificity than that of the MeSH main headings (also called descriptors). To improve the utility of MeSH for indexing the biomedical literature, in 1998 the Restrict to MeSH (RtM) algorithm was devised(2). The algorithm is designed to map entities from many other terminological systems to the MeSH thesaurus, regardless of whether the entities themselves are MeSH headings. Once entities from text are assigned MeSH headings, the semantic information they embody can be leveraged for a variety of purposes, including indexing.

The RtM method was evaluated several years ago on a limited basis but it has not yet been evaluated thoroughly. The goals of this project, therefore, were to conduct a qualitative and quantitative evaluation of the results of the algorithm and to recommend steps for its improvement. Although the high-level evaluation is essentially quantitative, it does convey some information about quality of the RtM output. The purpose of the qualitative evaluation is to assess the quality of the individual mappings.

## Background

Human indexing is expensive and labor-intensive(2). Given the exponential increase in the number of articles added to the biomedical literature, it is increasingly important to investigate automated indexing approaches. The goal of the NLM Indexing Initiative is to identify indexing approaches that can serve as a substitute, either completely or partially, for existing manual indexing approaches(2). The RtM method is a key component of the Initiative. It uses four basic approaches(3) to map a UMLS term to MeSH: through synonyms, through associated expressions, and through interconcept relationships, hierarchical and associative.

The Unified Medical Language System (UMLS) includes over 100 terminologies, classifications, ontologies, and other systems for representing biomedical knowledge. Together they are referred to as the UMLS source vocabularies. Although the vocabularies vary in form, function, and size, they are represented in the Metathesaurus in a standardized database format. A key unifying element of the Metathesaurus is a set of unique concepts, each of which is assigned a concept unique identifier, or CUI. Each concept in each of the source vocabularies is mapped to a single CUI. The 2006AA version of the Metathesaurus, which was used in this research, contains approximately 1.28 million concepts.

*MeSH*

MeSH includes three types of entities: main headings, qualifiers, and supplementary concepts. In this evaluation we chose to focus the analysis on main headings, which are central to MeSH indexing[1]. The main headings are organized into 16 polyhierarchical trees: Tree A contains anatomical terms, Tree B contains organisms, Tree C contains diseases, and so on. Some main headings are not intended for use in indexing. As illustrated in Figure 1, the majority of MeSH entities are either supplementary concepts or main headings.

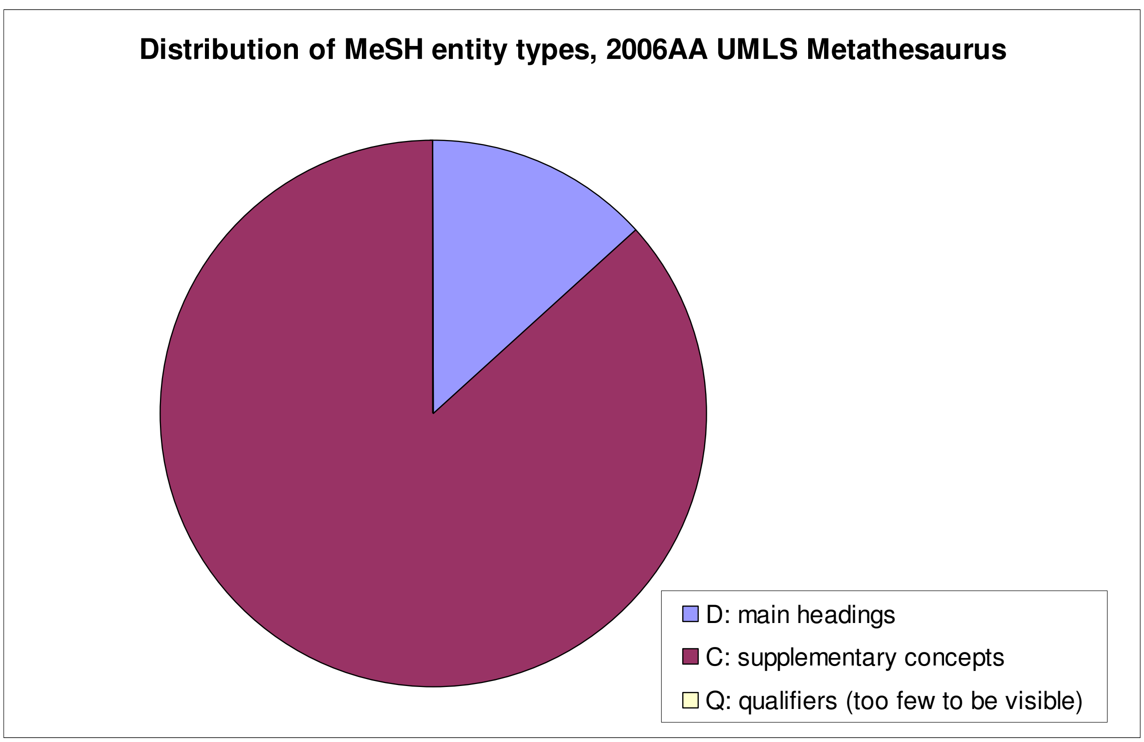


**Figure 1.** Distribution of MeSH entity types. The majority (156,957, 86.8%) of MeSH entities are supplementary concepts. Main headings, the focus of this evaluation (23,885,

[1] Every supplementary concept is associated with a main heading, and the RtM algorithm sometimes outputs main headings identified via these links. Mappings occurring in this way are treated equivalently in the analysis.

13.2%), are the main headings used for indexing. A small number of qualifiers (83, 0.05%) are also used in indexing, in combination with the main headings.

*RtM method*

The RtM method was developed mainly in the late 1990's. No significant changes have been made since that time[2]. NLM staff executes the algorithm once yearly on every CUI in the Metathesaurus. The results are available in a tool used by MEDLINE catalogers, the Data Creation and Maintenance System (DCMS)[3]. Some of the catalogers use the algorithm as a tool to support manual indexing; however, the method is not used for fully automated indexing.

The algorithm uses four approaches to map concepts to MeSH main headings(4):

1. Choose a MeSH term as a synonym of the initial concept.
2. Choose an associated expression which is the translation of the initial concept.
3. Select MeSH terms from concepts hierarchically related to the initial concept (this can only be attempted when a given CUI has at least one asserted relationship with an ancestor concept.)
4. Base the selection on the non-hierarchically related concepts of the initial concept.

These techniques rely on data from throughout the Metathesaurus, and each one is considered more aggressive than the previous. The first two are the most straightforward, while the last two often require many more steps, and although they can return appropriate mappings, they also carry the risk of suggesting less closely-related concepts due to semantic drift.

The first technique is to use synonymy. After a term is identified in text and represented as a CUI, the algorithm simply determines whether the CUI also corresponds to a term contributed by MeSH. The terms might be the same, or they might be lexically different, but if they are different, they are synonyms from the perspective of the UMLS. An example is the CUI C0002006 (Aldosterone). Aldosterone is also in MeSH (heading D000450), so the mapping is straightforward.

If the first approach is unsuccessful, the second is to use associated expressions, combinations of main headings and subheadings. The associated expressions are search strategies that have been created for representing terms from the International Classification of Diseases. For example, the term "Endoscopic removal of intraluminal foreign body from oesophagus without incision" can be mapped using this approach to the MeSH headings *Esophagus,* with the subheading *surgery*, along with the main heading *foreign bodies*.

If the first two approaches both fail, the third is to try to use the concept's ancestors. To do this, the algorithm looks at concepts that are related to the original concept via *parent-of* and *broader-than* relationships, as well the parents and broader

---

[2] In 2001, there was a subtle change in the algorithm. At this time, semantic groups were introduced as a way to prevent semantic drift when terms are mapped via the graph of ancestors.

[3] Although the suggested headings are sometimes used as a starting place for junior MEDLINE catalogers, subjective evidence indicates the suggested headings are most useful for less senior catalogers. More senior catalogers who have achieved a facility with MeSH have not depended on the mappings as a primary indexing tool.

concepts of these, and so on, all the way to the top of the graph. To prevent semantic drift, ancestors with a semantic type considered incompatible are not used. From the graph of all these ancestors, any concepts that come from MeSH are selected. To ensure that the most specific terms are selected, MeSH concepts that are ancestors of other MeSH concepts are removed. The algorithm also tries using the concept's children or siblings as a seed to build the graph of ancestors. As an example, the CUI *giant cell sarcoma* can be mapped to *sarcoma* by this method.

If the first three approaches are all unsuccessful, the algorithm attempts to find a mapping using what are called "other related concepts", which are concepts that have a relationship with the original concept via an "other relationship" link. In most cases, these correspond to nonhierarchical relations (also called associative relations). Using this method, the CUI "Nicotinic Acid 0.15 MG / Riboflavin 0.02 MG / Thiamine 0.06 MG Oral Tablet" can be mapped to the main headings "Niacin", "Riboflavin", "Thiamine", and "Tablets".

**Materials**

To prepare for the analysis we first obtained a data table containing a complete list of CUIs from the 2006AA UMLS Metathesaurus along with the corresponding MeSH headings suggested by the most recent run of the algorithm.

Figure 2 contains data associated with one row of this table. In this case, the concept[4] "Cervical nerve root compression" is paired with suggested mappings to the MeSH main headings "Neck Injuries", "Spinal Cord Compression", "Mononeuropathies", and "Radiculopathy".

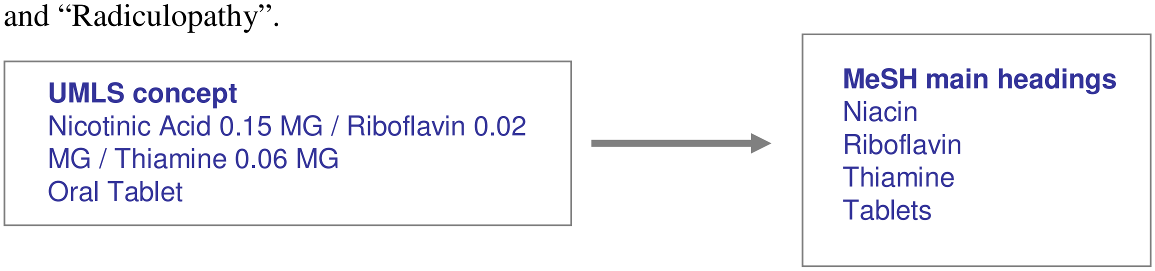


**Figure 2.** The data table used for the analysis contained one record for every CUI along with the associated suggested mappings to MeSH entities, if any. The example above is that of a concept that was mapped to four MeSH main headings by way of the graph of ancestors (seeded by parents).

We next obtained the 2006AA version of the Metathesaurus in Rich Release Format (RRF). We loaded the tables into a series of relational databases, which were then linked as needed.

**Methods**

*Quantitative evaluation*

The quantitative analysis was framed from three different perspectives: the perspectives of MeSH entities, CUIs, and mapping method used. We used SQL queries to compute statistical parameters relevant to the evaluation.

[4] The string used to represent the concept is its preferred form in English

*Qualitative evaluation*

To perform the qualitative evaluation, we first obtained a list of all CUIs. After excluding the CUIs for which mapping was accomplished via the straightforward *synonomy* method, we randomly selected 50 CUIs. We next confirmed that the selection roughly matched the distribution in mapping methods used by the algorithm for all CUIs. Finally, for each CUI and for each main heading suggested, we collected data to answer the following questions:

- What mapping method is used?
- If there is no mapping to MeSH, why not?
- Is the mapping in the same semantic neighborhood as the CUI? If not, how did this lateral semantic drift occur?
- Is the mapping at the same level of specification as the CUI? If not, how did this semantic drift occur?
- If the graph of ancestors was used, how many tree levels were climbed before selecting the suggested mapping?
- For CUIs that mapped to several MeSH entities, what proportion were appropriate mappings (none, some, or all)?

To answer these questions we relied in part on detailed output of the RTM mapping, available through a web interface(5).

## Results

The results of the analysis are summarized below. The results of the quantitative evaluation are discussed first, divided into the different perspectives, followed by the results of the qualitative evaluation.

### *From the perspective of MeSH entities*

We will first discuss the results from the perspective of the MeSH entities.

*As expected, the majority of MeSH entities suggested were main headings*

As illustrated in Figure 3, among the three MeSH entity types, the majority of mappings[5] (82.8%) were to main headings. Among the remaining MeSH entities suggested, 13.9% were supplementary concepts, and 3.3% were qualifiers. When these results are compared with the overall distribution of the MeSH entity types (Figure 1), it is clear that the algorithm suggested proportionately more main headings than supplementary concepts. Because the main headings are of foremost importance for use in indexing, they are the focus of this evaluation.

[5] The denominator for the proportions is the total number of MeSH tokens occurring as suggested mappings, rather than the number of unique main headings, subheadings, and supplementary concepts.

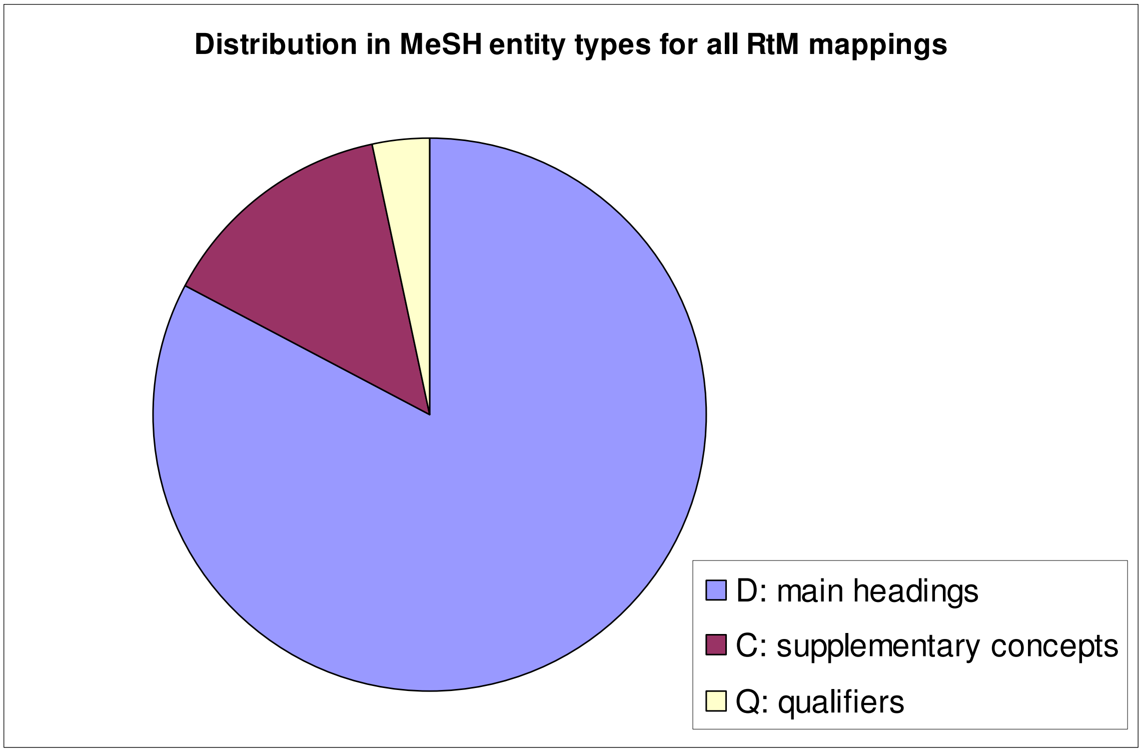


**Figure 3.** Distribution in MeSH entity types for all RtM mappings. Among the three MeSH entity types, the majority of mappings (82.8%) were to main headings. Among the remaining MeSH entities suggested, 13.9% were supplementary concepts and 3.3% were qualifiers. The main headings are the MeSH entities which are mainly used for indexing, and are the focus of the evaluation.

*Distribution in tree depth of suggested main headings was similar to distribution of main headings in all of MeSH*

An analysis of the depth of the main headings suggested by RtM demonstrates that the algorithm mapped CUIs to terms throughout the MeSH hierarchies (Figure 4). As with the distribution in tree depths of all MeSH main headings (Figure 5), the most common tree depth of the suggested mappings was four. When compared with the distribution in depth of all main headings, the distribution in suggested mappings indicates that the algorithm preferentially suggested more general terms (terms on levels one and two of the hierarchies.) Main headings on the first two levels comprised 13.6% of suggested mappings, but such main headings represent just 6.5% of all the main headings in MeSH.

Many of the main headings which are not intended for use in indexing are situated in the top two levels of the hierarchies. However, since multiple main headings were suggested for many CUIs, only 26,971 (2.1%) of the CUIs were mapped such that the only main heading suggested was a main heading not used for indexing[6].

[6] There were 36,680 CUIs that mapped to at least one main heading which is not used for indexing. Fifty-six CUIs mapped to exactly three such main headings and 2,486 CUIs mapped to exactly two such main headings. 34,138 CUIs mapped to just one such main

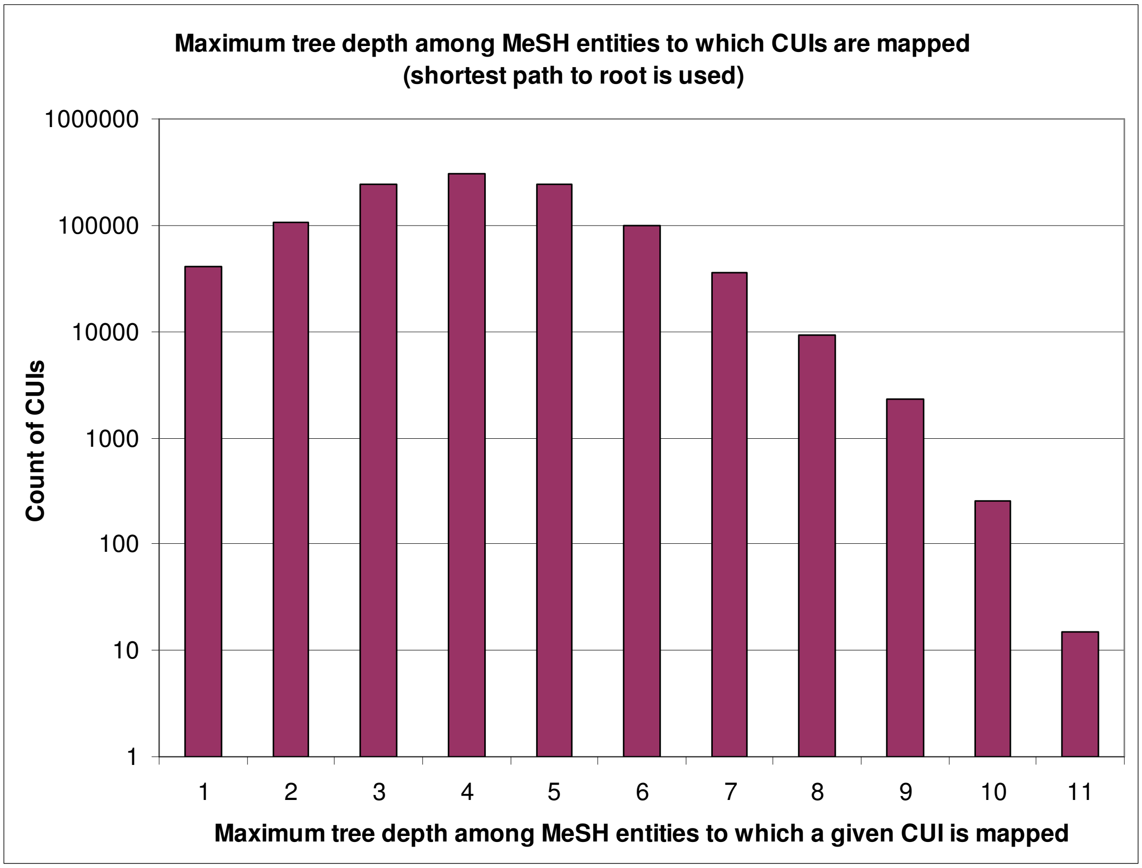


**Figure 4.** Maximum tree depth among MeSH entities to which CUIs are mapped. The RtM algorithm selected concepts at all tree depths throughout the MeSH hierarchies. The most common tree depth of the selected mappings was four. When compared with the distribution in depth of all main headings, the distribution in suggested mappings indicates that the algorithm preferentially suggested more general terms (terms on levels one and two of the hierarchies.)

---

heading. Among the 36,680 were 26,971 CUIs for which the only main heading or main headings suggested were among the main headings not used for indexing.

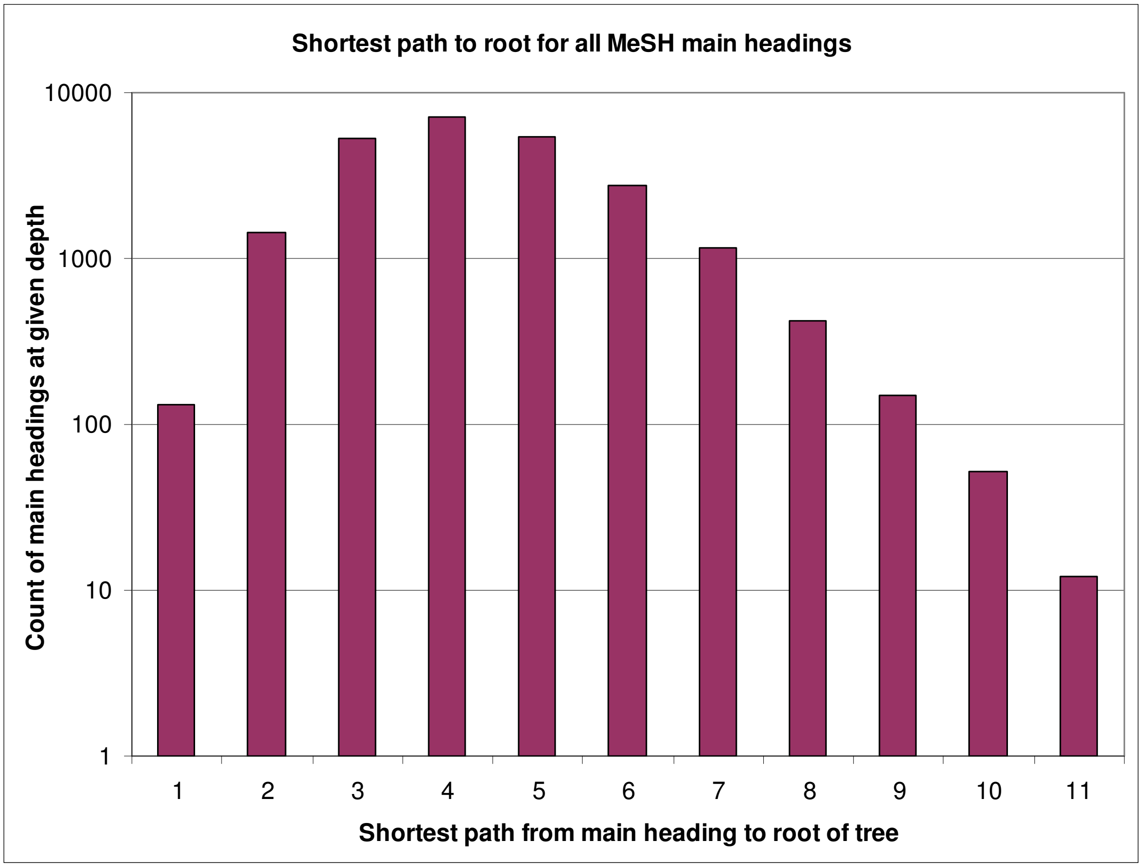


**Figure 5.** Shortest path to root for all MeSH main headings (for comparison against Figure 4). Main headings are situated throughout the MeSH trees. As in Figure 4, the mode of the distribution is a depth of four.

To analyze the MeSH tree depths in more detail, we divided the MeSH trees into five depth levels. We calculated the shortest distance from the root to every concept in MeSH, and then assigned each CUI to a depth category from one to five. For example, in MeSH tree A (anatomy), the concept “Bronchial tree” is three levels down. Tree A, in all, has 10 levels. Thus “Bronchial tree” was assigned a depth category of 2, which is near the root. Figure 6, based on data calculated using this method, includes the distribution of main headings in all of MeSH, as well as for MeSH main headings to which some mapping was provided. This distribution for all of MeSH is skewed to the right, while the distribution for suggested main headings is not skewed. This indicates that RtM preferentially suggested headings that were higher in the trees. This result is not surprising since the algorithm often relies on the mapping method in which the graph of ancestors is traversed to identify more general concepts. Overall these results are encouraging since the basic distribution is similar to that for all of MeSH.

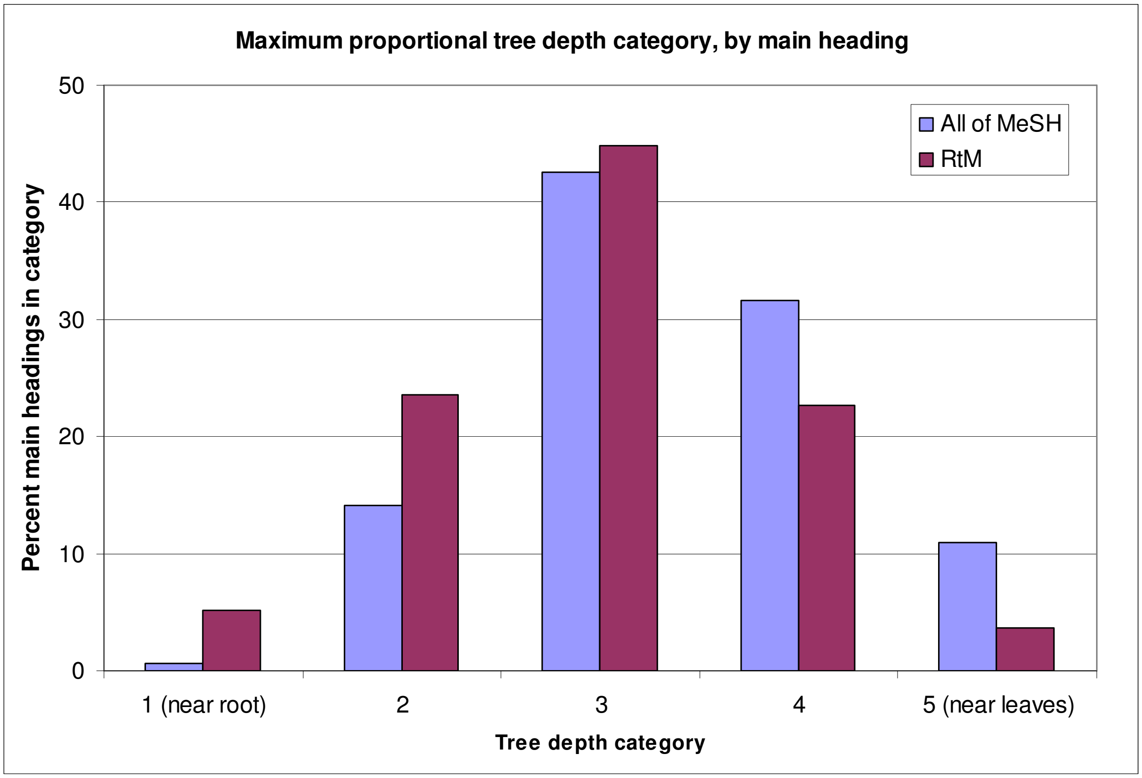


**Figure 6.** Maximum proportional tree depth category, by main heading. This histogram shows how many main headings exist at various depths in the MeSH trees, alongside the proportion of mappings at that level of depth. Overall these results are encouraging since the basic distribution is similar to that for all of MeSH.

*The algorithm suggested main headings in all MeSH trees*

As shown in Figure 7, the algorithm suggested main headings in all 16 MeSH trees. For all but three trees [K (Humanities), V (Publication Characteristics), and Z (Geographic Locations)], there were mappings to over half of the possible main headings in each tree. These three trees have many concepts that can be reached only by synonymy, and not via links from other terminologies. This indicates that MeSH coverage in these three areas is rich when compared with the coverage of other UMLS source vocabularies.

Figure 8 conveys the total number of main headings in each tree as well as the proportion that were mapped via a method other than the straightforward *synonymy* method. For all of MeSH, relatively few main headings were mapped via synonymy; 73.2% were mapped using one of the more aggressive methods. For geographic locations (Tree Z), only a small proportion of CUIs were suggested for any reason other than the fact that they were contributed by MeSH. By contrast, for anatomical concepts (Tree A), the converse was true: A high proportion of headings in tree A could be reached via vocabularies other than MeSH. As the figure shows, RtM mapped to a large proportion of the MeSH headings in each tree.

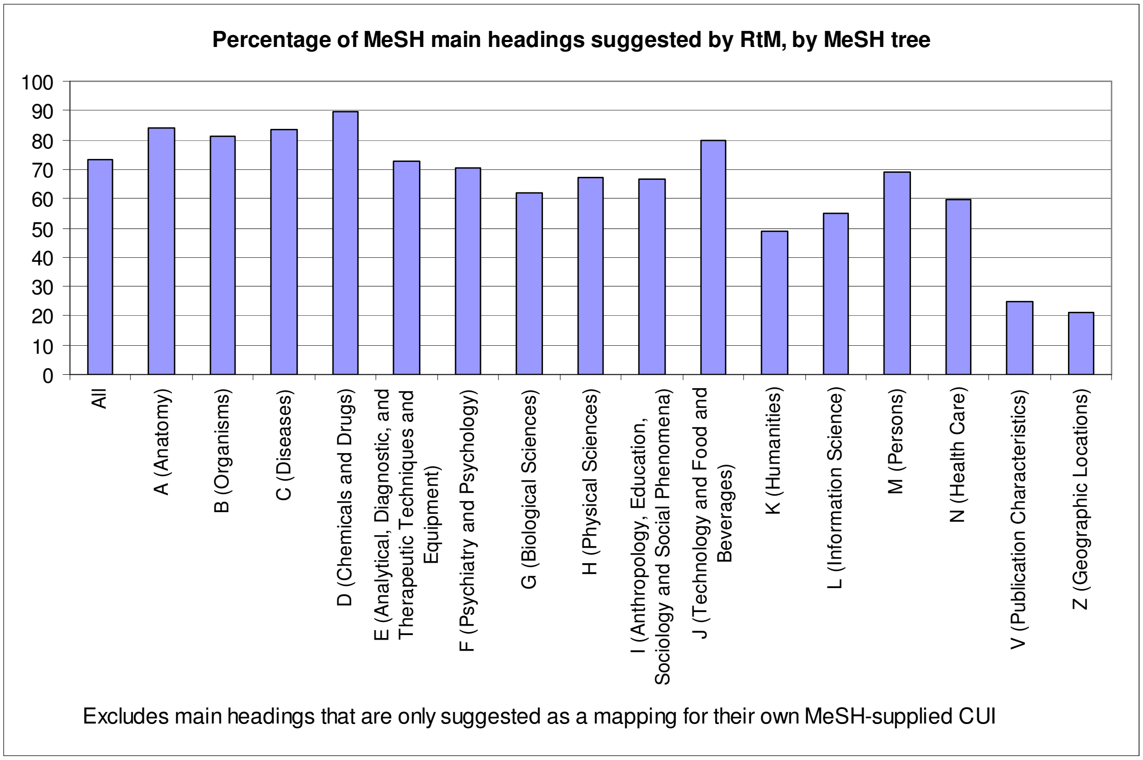


**Figure 7**. Percentage of MeSH main headings suggested by RtM, by MeSH tree. For all but three trees [K (Humanities), V (Publication Characteristics), and Z (Geographic Locations)], there were mappings to over half of the possible main headings in each tree. These three trees have many concepts that can be reached only by synonymy, and not via links from other terminologies. This implies that MeSH coverage in these three areas is rich when compared with the coverage of other UMLS source vocabularies.

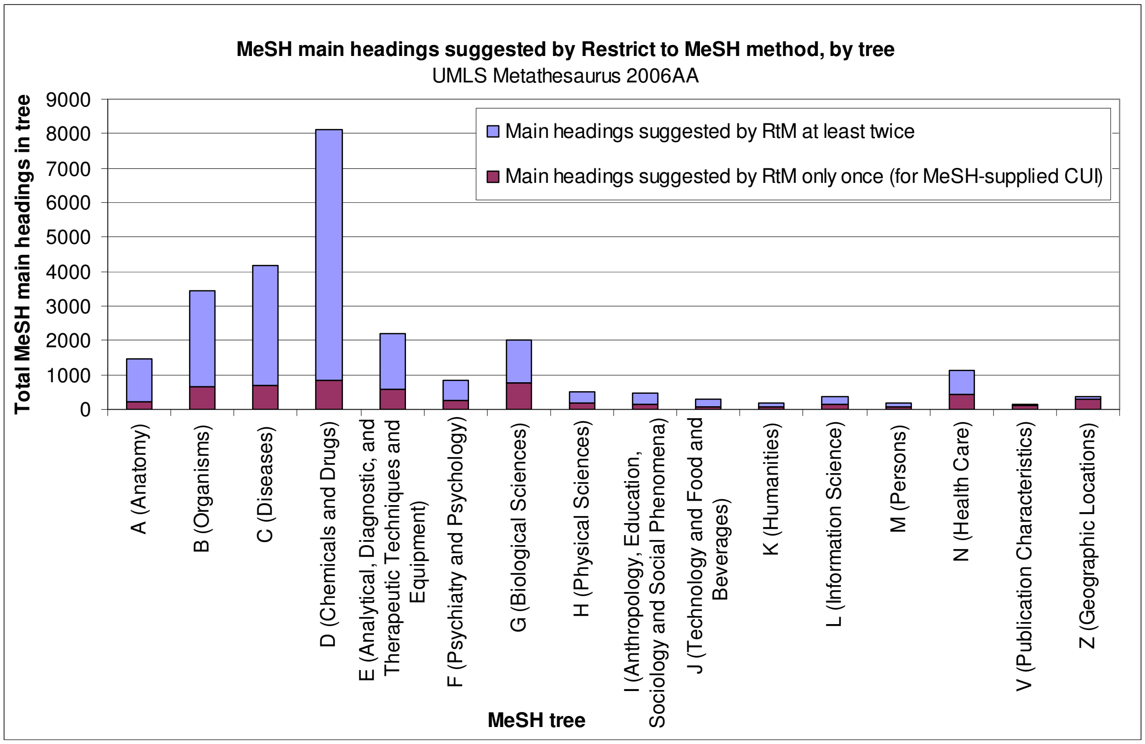


**Figure 8.** MeSH main headings suggested by RtM, by MeSH tree. This graph conveys the total number of main headings in each tree as well as the proportion that were mapped via a method other than the straightforward *synonymy* method. For all of MeSH, relatively few main headings were mapped via synonymy; 73.2% were mapped using one of the more aggressive methods. For geographic locations (Tree Z), only a small proportion of CUIs were suggested for any reason other than the fact that they were contributed by MeSH. By contrast, for anatomical concepts (Tree A), the converse was true: A high proportion of headings in tree A could be reached via vocabularies other than MeSH. As the figure shows, RtM mapped to a large proportion of the MeSH headings in each tree.

***From the perspective of UMLS CUIs***

The following analyses were carried out from the perspective of UMLS CUIs.

*Most CUIs were assigned suggested mappings to at least one MeSH entity*

With respect to the algorithm's overall performance in mapping, 84.5 percent of CUIs were assigned at least one mapping to MeSH (Figure 9). Although it is not possible to discuss the quality of the mappings, it indicates that at least some MeSH entity was found for the majority of the CUIs.

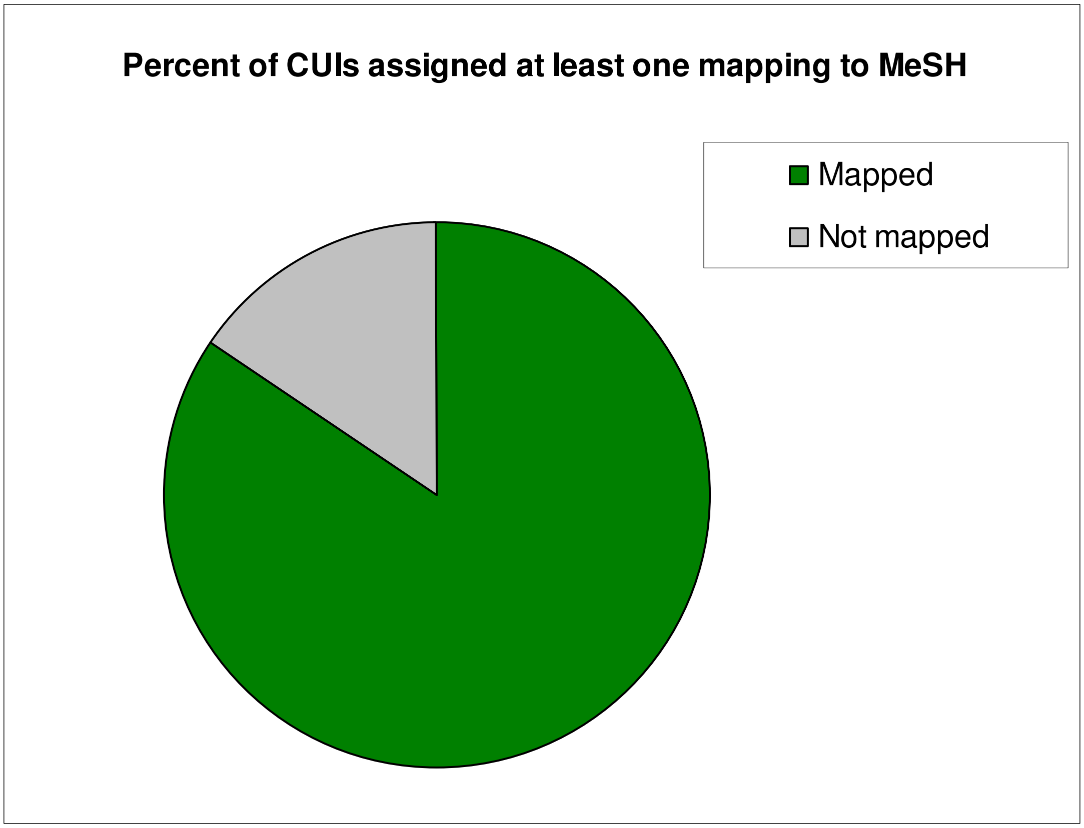


**Figure 9**. Percent of CUIs assigned at least one mapping to MeSH. 84.5 percent of CUIs were assigned at least one mapping. Although it is not possible to discuss the quality of the mappings, it indicates that at least some MeSH entity was found for the majority of the CUIs.

*Frequency of mappings to individual MeSH main headings was characterized by a power law distribution*

Mappings to MeSH main headings were distributed about a power-law distribution, appearing nearly linear when displayed on a double-logarithmic scale (Figure 10). This indicates that there were a small number of main headings to which a very large number of CUIs were mapped, and a very large number of main headings to which only a few CUIs were mapped. Among these main headings were general concepts such as D001829 (Body Regions), D013607 (Tablets), and D013502 (Surgery).

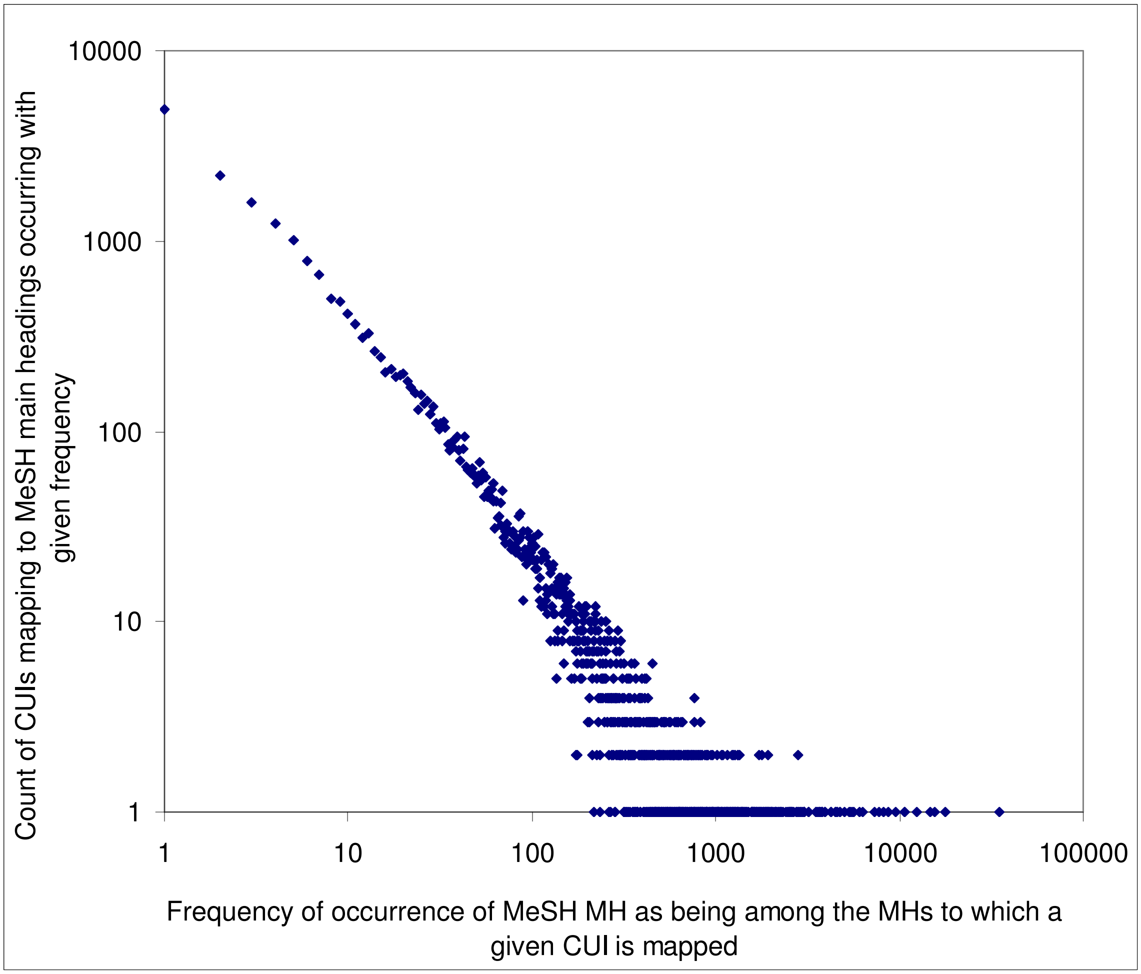


**Figure 10.** Frequency distribution of mappings to all MeSH main headings. Mappings to MeSH main headings were distributed about power-law distribution, appearing nearly linear when displayed on a double-logarithmic scale. This indicates that there were a small number of main headings to which a very large number of CUIs were mapped, and a very large number of main headings to which only a few CUIs were mapped.

*Percentage of CUIs mapped and mapping methods varied by source family*

The performance of the RtM method in mapping CUIs to MeSH main headings varied in part based on the source of the CUIs. Figure 11 indicates the proportion of CUIs contributed by each source family for which the method was able to suggest at least one main heading. Figure 12 shows that the mapping method also varied by source family. The source families are grouped by predominant mapping method used. For the majority of source families, the mapping method used most was the graph of ancestors. Notable exceptions include LNC (Logical Observation Identifier Names and Codes, Version 2.15.) and JABL (Online Congenital Multiple Anomaly/Mental Retardation Syndromes, 1999). For LNC, most mappings were identified via otherwise related concepts, while for JABL, most mappings were based on associated expressions. UMLS source vocabulary abbreviations and names are available online(6)

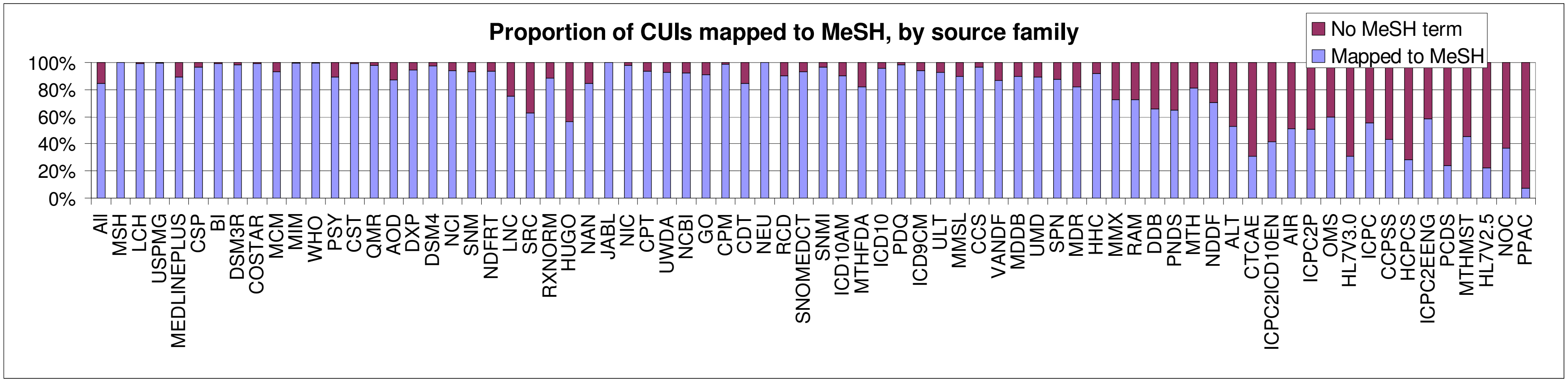


**Figure 11.** Proportion of CUIs mapped to MeSH, by source family. This histogram shows the proportion of CUIs contributed by each source family for which RtM was able to suggest at least one main heading.

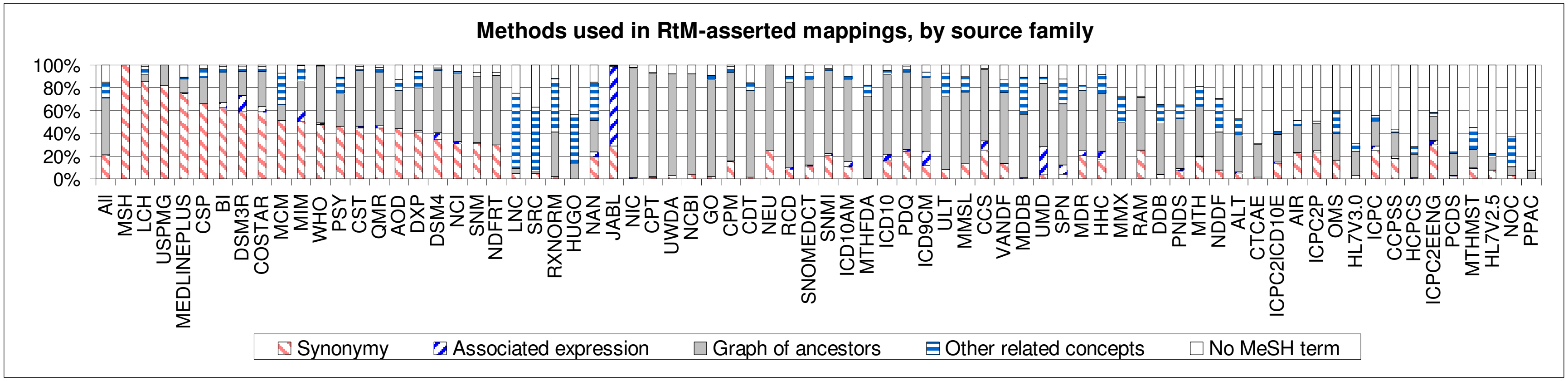


**Figure 12.** Methods used in RtM-asserted mappings by source family. This histogram shows the differences in mapping methods for different source families. The source families are grouped by predominant mapping method used. For the majority of source families, the mapping method used most was the graph of ancestors. Notable exceptions include LNC (Logical Observation Identifier Names and Codes, Version 2.15.) and JABL (Online Congenital Multiple Anomaly/Mental Retardation Syndromes, 1999). For LNC, most mappings were identified via otherwise related concepts, while for JABL, most mappings were based on associated expressions. UMLS source vocabulary abbreviations and names are available online(6).

Figure 13 contains the proportion of CUIs mapped in each semantic group, stratified by mapping method used. In this graph, *chemicals and drugs* were the semantic group for which the highest proportion of concepts mapped to MeSH. This is due to the fact that most of the concepts in the semantic group *chemicals and drugs* were mapped via synonymy, the most straightforward of the four methods. The reason for this is that many of the CUIs in the Metathesaurus which are chemicals and drugs are also among the approximately 150,000 supplementary concepts in MeSH. By contrast, for disorders, there were relatively few mappings via synonymy. This is because the *disease* tree in MeSH includes only about 8,000 main headings, whereas among all CUIs there are around 213,000 which are in the disorders semantic group. For the disorders, most of the mappings were done via the graph of ancestors, seeded by parents, which means that disease terms in most terminologies are finer-grained than in MeSH. The same is true for *anatomical concepts*, *living beings*, *procedures*, *devices*, and *genes and molecular sequences*. This reflects the fact that there are multiple terminologies in the UMLS that contain lists of specific concepts in these areas.

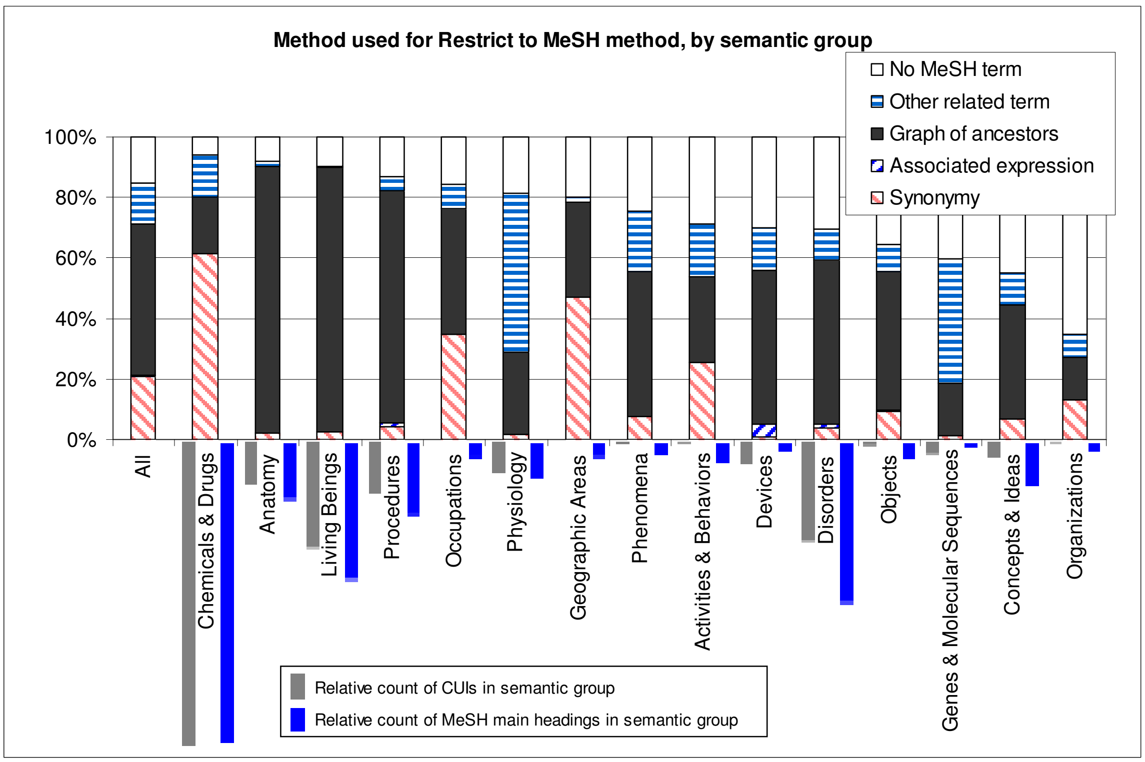


**Figure 13.** Methods used for RtM method, by semantic group. For each semantic group there are two bars originating at the x-axis and pointed downward. The bar on the left indicates the number of CUIs in each semantic group relative to the 650,000 in Chemicals & Drugs. The bar on the right represents the number of MeSH main headings in each semantic group relative to the approximately 8,000 in Chemicals & Drugs.

As discussed earlier, one approach used by the algorithm is to assign mappings based on ancestors. By definition, this method cannot be used on CUIs that have no ancestor contributed by any source. Figure 14 contains the count and proportion of such CUIs contributed by each source family. Most CUIs with no ancestor were still possible to map to MeSH, often via other related concepts.

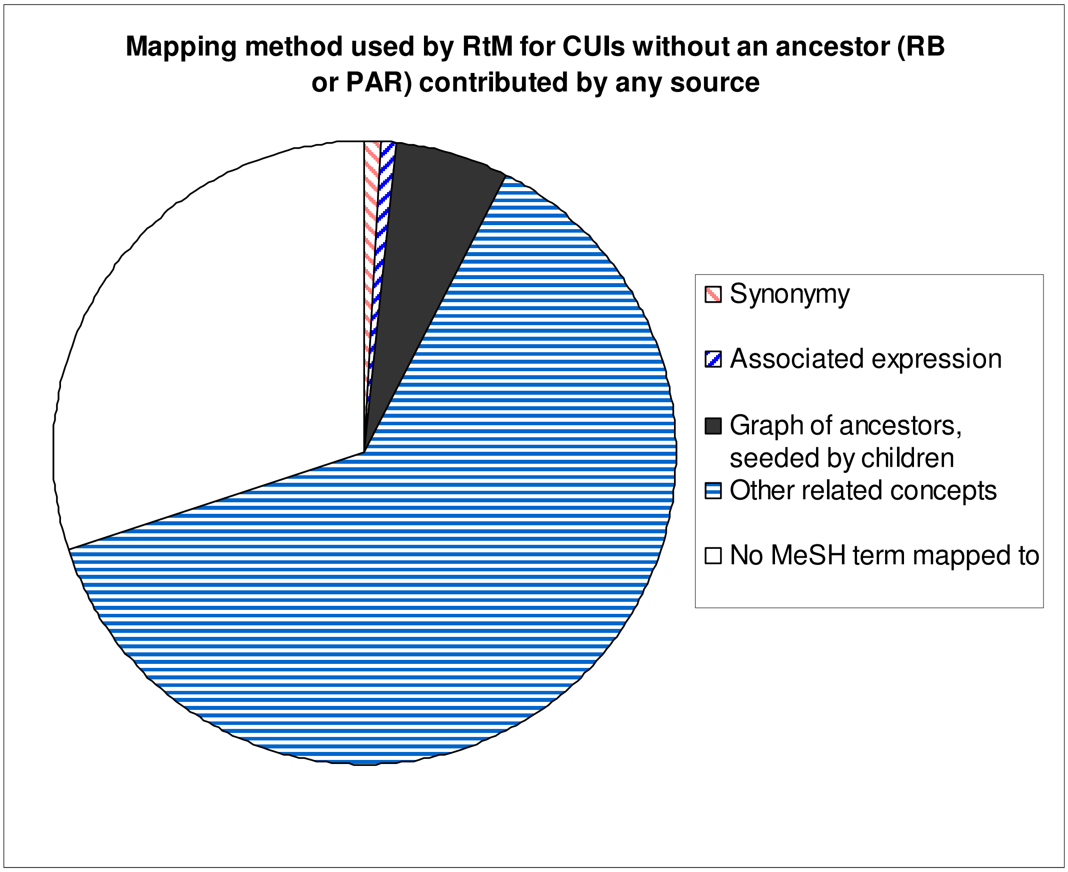


**Figure 14.** Mapping methods used for CUIs having no ancestor based on asserted parent-of or broader-than relationship. Most CUIs with no ancestor were still possible to map to MeSH, often via *other related concepts.*

Figure 15 depicts the mapping method used in each proportional tree category of suggested mappings. This histogram shows that CUIs that mapped to more specific concepts (in categories 4 and 5 above) were more likely than more general CUIs to be mapped via associated expressions and other related concepts, and less likely than the more general CUIs to be mapped via the graph of ancestors.

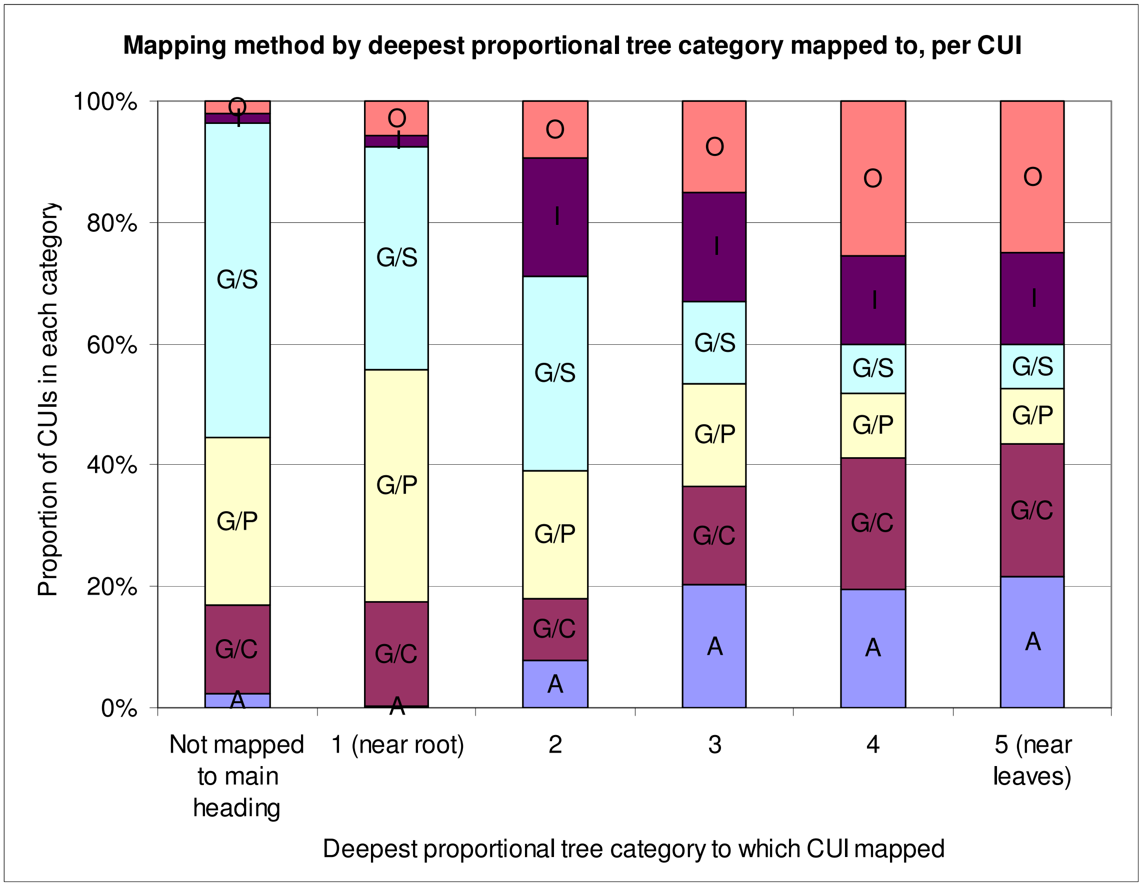


**Figure 15**. Mapping method by proportional tree category of suggested mappings. This histogram shows that CUIs that mapped to more specific concepts (in categories 4 and 5 above) were more likely than more general CUIs to be mapped via associated expressions and other related concepts, and less likely than the more general CUIs to be mapped via the graph of ancestors.

*A small number of MeSH-supplied CUIs were not mapped*

Finally, a small number of CUIs were not mapped to MeSH despite the fact that MeSH was among the source vocabularies contributing them. Ten of these contained special characters that were not compatible with the algorithm, while the remaining 22 of them were references to MeSH constructs such as “Organisms (MeSH Category)” and “MeSH Qualifiers”.

***From the perspective of mapping method***

The final perspective we use the quantitative analysis is that of mapping method. As seen in Figure 16, roughly half of the CUIs were mapped via the graph of ancestors.

Among the mappings via the graph of ancestors, the vast majority of the mappings were accomplished by using the concept's parents as the seed. The next most common methods were via *synonomy*, and via *other related concepts*. A small proportion of mappings were accomplished using *associated expressions*.

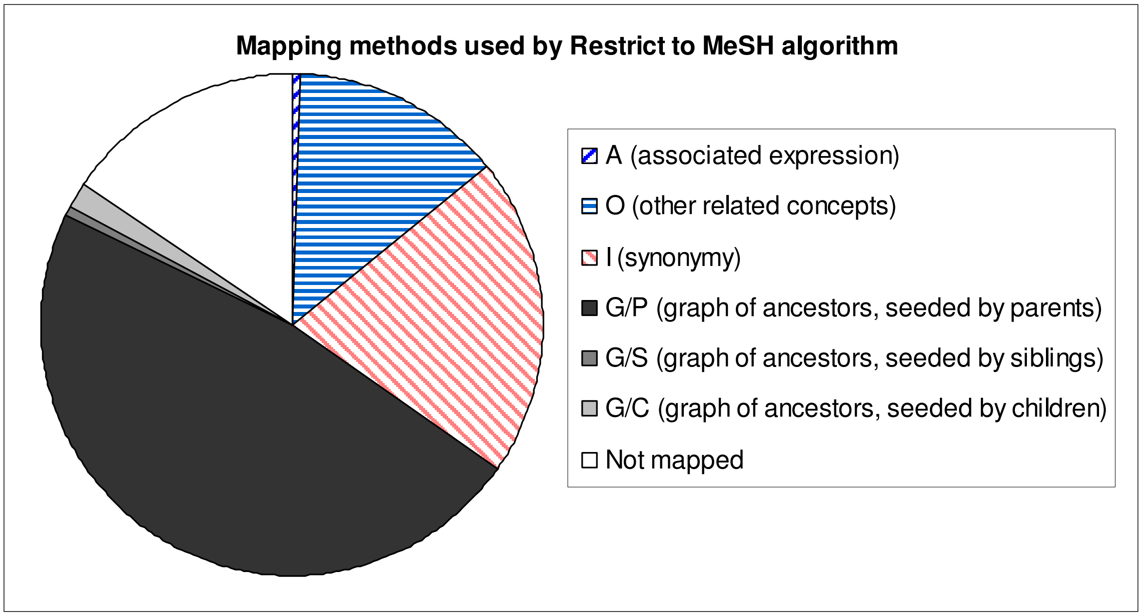


**Figure 16.** Mapping methods used by Restrict to MeSH algorithm. Roughly half of the CUIs were mapped via the graph of ancestors. Among the mappings via the graph of ancestors, the vast majority of the mappings were accomplished by using the concept's parents as the seed. The next most common methods were via *synonomy*, and via *other related concepts*. A small proportion of mappings were accomplished using *associated expressions*.

***Preliminary results of qualitative evaluation***

After drawing the random sample of 50 CUIs, we began by verifying that the mapping methods used for the random sample matched those used for all CUIs, and they were indeed similar (Figure 17).

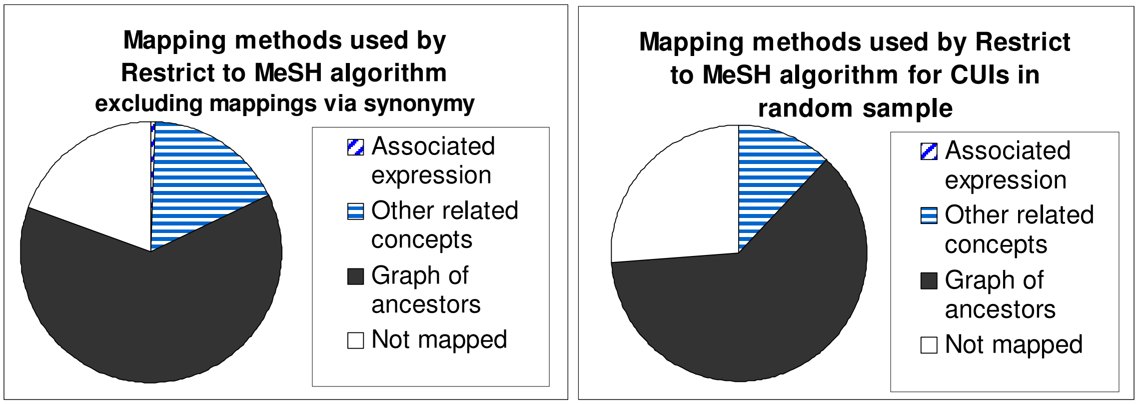


**Figure 17.** Methods used by Restrict to MeSH algorithm for all CUIs and for 50 CUIs in random sample. Comparing the two graphs shows that the mapping methods used for the CUIs in the random sample were roughly similar to the methods used for all CUIs.

As expected, there were several CUIs (n=13) that were not assigned any mapping to MeSH. The number of such CUIs was higher than expected based on the overall results. This is partly due to chance; there were a slightly higher number of such CUIs in the random sample. The number is also higher because the denominator excluded CUIs mapped via synonymy. Of the 13 that weren't mapped, six could not be mapped because not only were mappings via *synonymy* and *other related concepts* impossible, but mapping via the graph of ancestors could not be attempted because the concepts were orphans (CUIs with no parent or broader concept contributed by any source vocabulary). Examples of concepts that fit both of these conditions are "Referral, allied health prof" and "Fear (of) surgery" which are both orphan concepts contributed by ICPC2P, and "C6orf86 gene" which is an orphan concept contributed both by HUGO and by MTH.

Of the remaining seven that were not mapped, six were not orphans but could not be mapped because traversing the graph of ancestors only identified other concepts which were themselves unmappable. One example of such a concept is "Vagal gastric function", which is contributed by the SNOMED family. Walking up the tree only identifies other entities which are also contributed by SNOMED but which cannont be assigned mappings to MeSH.

For the remaining CUI among the 13 that could not be mapped, the problem was that although a parent concept was identified, it was disregarded by the algorithm because it was not in the same semantic category as the original CUI. An example is "Visual acuity tests normal" which was contributed by the MedDRA family.

Among the 37 CUIs that were mapped, 89% of the main headings suggested were in the same semantic neighborhood as the original CUI. Of the four that were not, the problem was that there was a semantic boundary crossed in a hierarchical relationship in one of the vocabularies. For example, in the Read Codes, one of the concepts involving an adverse reaction to a drug was the child of a concept that was itself a drug[7]. All 37 were more general than the original CUI. This result was not surprising, given the fact that the algorithm often suggests mappings to MeSH main headings that are more general than the original CUI.

**Conclusion**

The results of this evaluation support the notion that the RtM method already achieves a fairly high level of performance in mapping concepts to MeSH. These are preliminary results which will be refined further in coming months. One possible outcome will be a paper for the 2006 MedInfo conference, which will possibly include a set of recommendations for improving the method. In conclusion, as the growth in biomedical knowledge continues to accelerate, further development of RtM is a key next step towards more effective manual and automated indexing methods.

**References**

1. McCray AT, Ide NC. Design and Implementation of a National Clinical Trials Registry. J Am Med Inform Assoc. 2000 May 1, 2000;7(3):313-23.
2. Aronson AR, Bodenreider O, Chang HF, Humphrey SM, Mork JG, Nelson SJ, et al. The NLM Indexing Initiative. Proceedings / AMIA. 2000;Annual Symposium.:17-21.
3. Bodenreider O, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. Proceedings / AMIA. 1998;Annual Symposium.:815-9.
4. Bodenreider O. Using UMLS semantics for classification purposes. Proceedings / AMIA. 2000;Annual Symposium.:86-90.
5. Bodenreider O. Restrict to MeSH web interface. National Institutes of Health. Available from URL: http://mor.nlm.nih.gov/perl/rtm.pl; 1998.
6. National Library of Medicine. 2006AA Section B.4. Source Vocabularies (UMLS). Available from URL: http://www.nlm.nih.gov/research/umls/archive/2006AA/metab4.html#sb4_0; 2006.

[7] In the Read codes, "Adverse reaction to drugs primarily affecting skin and mucous membrane, opthalmological, otorhinolaryngological and dental drugs" is related to the concept "Drugs, medicines and biological substances causing adverse effects in therapeutic use", via a parent/child relationship.